90-50J

From the last claim of Theorem 4.1, we can assume that $b/a < \beta$, which implies that $a > 1/\beta$ since $b > 1$. Furthermore, $a' < 1/\alpha$ implies that $a < \frac{1}{2}(1 + 1/\sqrt{\beta})$. Hence, the question is: can the function $f(x)$ be greater than $1/\alpha$ for all $x \in (1/\beta, \frac{1}{2}(1 + 1/\sqrt{\beta}))$. Now a very messy calculation will show that $f'(x) < 0$ on this interval. Alternatively, one can see that with b, β fixed and a increasing, $\mathcal{R}(a, b) = f(a)$ can only decrease, this being an optimum constrained sensitivity which can be made smaller the smaller is this constraint. This argument shows that $f'(x) \leq 0$ throughout the interval. Quite evidently, the explicit form of $f(x)$, being analytic in x, prevents $f'(x)$ being identically zero on any subinterval. Hence, in the interval $[1/\beta, \frac{1}{2}(1 + 1/\sqrt{\beta})]$, $f(x)$ attains its minimum only at the right endpoint (in fact, it can be checked that this endpoint is a uniquely possible extreme point of $f(x)$ in its domain of definition). An easy calculation shows that at the endpoint, $f[\frac{1}{2}(1 + 1/\sqrt{\beta})] = 1/\alpha$. Thus, $f(x) > 1/\alpha$ for all interior points in the interval and in particular $f(a) > 1/\alpha$.

Case 2: Suppose that $b' < 1/\alpha$ and $b^2 - b < (2ab - a^2 - b)\beta$. Then arguing as for Case 1, we can show that

$$\frac{b + \sqrt{b'\beta(\beta - 1)}}{\beta - b} > 1/\alpha.$$

Case 3: Suppose that the first three alternatives of Theorem 4.1 are precluded. We must show that if $b/a < \beta$, then $h(a, b) > 1/\alpha$, where

$$h(x, y) = \sqrt{\frac{x}{1-x}\frac{y}{y-1}\left(\frac{\beta - 1}{x\beta - y} - 1\right)}.$$

First, it is not hard to see that there are only the following three possibilities for a and b:

$$a' < 1/\alpha < b' \quad \text{and} \quad a + b^2 - 2ab \geq (a - a^2)\beta \qquad (A.6)$$

$$b' < 1/\alpha < a' \quad \text{and} \quad b^2 - b \geq (2ab - a^2 - b)\beta \qquad (A.7)$$

$$\max(a', b') \leq 1/\alpha. \qquad (A.8)$$

If a and b satisfy (A.6), it was shown in the proof of Theorem 4.1 that

$$h(a, b) \geq b' > 1/\alpha.$$

In the same way, it can be shown that if a and b satisfy (A.7), then

$$h(a, b) \geq a' > 1/\alpha.$$

Now suppose that a and b satisfy (A.8), or equivalently,

$$a \leq x_0 \triangleq \frac{1}{2}(1 + 1/\sqrt{\beta}) \quad \text{and} \quad b \geq y_0 \triangleq \frac{1}{2}(1 + \sqrt{\beta}).$$

Since $\min(a', b') < 1/\alpha$, the above two inequalities cannot be replaced by equalities simultaneously. Since $\mathcal{R}(a, b) = h(a, b)$ is a constrained optimum, it is intuitively clear that

$$\frac{\partial}{\partial x}h(x, y) \leq 0 \quad \text{and} \quad \frac{\partial}{\partial y}h(x, y) \geq 0.$$

In addition, neither partial derivative can be identical to zero on an interval. As a direct calculation shows, $h(x_0, y_0) = 1/\alpha$. Since either $a < x_0$, $b \geq y_0$, or $a \leq x_0$, $b > y_0$, it follows that

$$h(a, b) > 1/\alpha.$$

Finally, Theorem 4.2 is concluded by combining the above arguments. □

References

[1] J. C. Doyle and G. Stein, "Multivariable feedback design: Concepts for a classic/modern synthesis," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 4-16, 1981.

[2] J. S. Freudenberg and D. P. Looze, "Right half plane poles and zeros and design tradeoffs in feedback systems," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 555-565, 1985.

[3] B. A. Francis and G. Zames, "On H^∞-optimal sensitivity theory for SISO feedback systems," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 9-16, 1984.

[4] P. P. Khargonekar and A. Tannenbaum, "Non-Euclidean metrics and the robust stabilization of systems with parameter uncertainty," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 1005-1013, 1985.

[5] H. Kimura, "Directional interpolation approach to H^∞-optimization and robust stabilization," *IEEE Trans. Automat. Contr.*, vol. AC-32, pp. 1085-1093, 1987.

[6] A. Tannenbaum, "Feedback stabilization of plants with uncertainty in the gain factor," *Int. J. Contr.*, vol. 32, pp. 1-16, 1980.

[7] —, *Invariance and System Theory: Algebraic and Geometric Aspects.* New York: Springer-Verlag, 1981.

[8] D. C. Youla, J. J. Bongiorno, and C. N. Lu, "Single-loop feedback-stabilization of linear multivariable dynamical plants," *Automatica*, vol. 10, pp. 159-172, 1974.

[9] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 301-320, 1981.

[10] G. Zames and B. A. Francis, "Feedback and minimal sensitivity and optimal feedback," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 585-601, 1983.

[11] H. Kober, *Dictionary of Conformal Representations.* New York: Dover, 1952.

[12] J. H. Curtiss, *Introduction to Functions of a Complex Variable.* New York: Marcel Dekker, 1978.

[13] J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, vol. 20. AMS Colloquium Publications, 1965.

A Unifying Theorem for Linear and Total Linear Least Squares

BART DE MOOR AND JOOS VANDEWALLE

Abstract—**It is shown how both linear least squares and total linear least squares estimation schemes are special cases of a rank one modification of the data matrix or the sample covariance matrix. For a problem with n unknowns, there exist n linear least squares solution while the total linear least squares solution is (generically) unique. When the signal-to-noise ratio is sufficiently high, the total least squares solution is a nonnegative combination of the least squares solutions.**

I. Introduction

Among the most popular schemes for estimating linear relations from noisy data are the Linear Least Squares (LLS) and the Total Linear Least Squares (TLLS) schemes. The literature on LLS is vast and the problem has a long history, starting with Gauss and Legendre. It is most commonly used in signal processing and system identification because of its straightforward geometrical interpretation (the orthogonality principle), its structure, (which is optimally suited for recursive implementations), and the relative ease by which statistical and numerical properties can be derived. References [7], [8], and [10]–[12] provide good surveys on the numerical and statistical richness of the subject. The TLLS problem, known in the statistical community as "orthogonal regression," can be traced back over more than 100 years, being rediscovered many times [1]. The term TLLS was coined in [6], although its solution, using the singular value decomposition, was introduced in [5]. Statistical properties, algorithms, and applications in signal processing and system identification can be found in [4] and [13].

Manuscript received July 27, 1987; revised December 27, 1988. Paper recommended by Past Associate Editor, H. L. Weinert. The work of B. De Moor was supported by grants from NATO and IBM.

B. De Moor is with the Department of Electrical Engineering, ESAT Laboratory, Katholieke Universiteit Leuven, Kardinal Mercierlaan 94, 3030 Heverlee, Belgium, on leave at the Department of Computer Science and the Information Systems Lab, Stanford University, Stanford, CA 94305.

J. Vandewalle is with the Department of Electrical Engineering, ESAT Laboratory, Katholieke Universiteit Leuven, Kardinal Mercierlaan 94, 3030 Heverlee, Belgium.

IEEE Log Number 9034490.

0018-9286/90/0500-0563$01.00

The problem of identifying linear relations from noisy data can be cast in a matrix formulation as follows.

Let the $m \times n$ real matrix A with $m > n$ represent m different measurements on n measured channels. At least one linear relation exists between the columns of the matrix A if there is a nonzero vector x such that

$$Ax = 0. \tag{1}$$

This is equivalent to the algebraic condition:

$$r = \text{rank}(A) < n$$

indicating that there exist $n - r$ linearly independent linear relations between the columns of A. Generically, however, real measurements will always be such that A is of full column rank n. Most frequently, one then makes the assumption that the data were generated by "exact" data that have been perturbed by noise (uncertainties, inaccuracies, etc.) in an *additive* way. Denote by $\hat{A}$ the exact and by $\tilde{A}$ the noise matrix. Then

$$A = \hat{A} + \tilde{A}.$$

The problem of identifying linear relations from the data matrix A can be considered as modifying A to obtain a rank deficient matrix $\hat{A}$, based upon several *a priori* assumptions to be made (recently, some of these *a priorisms* have received a good deal of criticism, as discussed in [2], [3], and [9]). Once a rank deficient matrix $\hat{A}$ has been obtained, the linear relations follow immediately from the kernel of $\hat{A}$ as in (1).

We will make the following assumptions throughout the note.

- The noise is additive and *all* columns of the data matrix A are perturbed by noise.
- $r = \text{rank}(\hat{A}) = n - 1$. This implies that we are looking only for 1 linear relation. This can be achieved by sufficiently high signal-to-noise ratios in static estimation, by reducing the number of variables from statistical dependency tests, or by assuring that conditions of persistant excitation are satisfied in a dynamical estimation scheme. Of course, the precise estimation of r itself given only the raw data is a highly nontrivial task which is, however, not discussed here.

Both LLS and TLLS schemes modify the available data matrix A in a particular way, which is explored in Sections II and III. In Section IV, it will be shown how both LLS and TLLS are special cases of a rank one modification of the data matrix A or the sample covariance matrix $\Sigma = A^t A$.

II. Linear Least Squares

Denoting by a_i the ith column of A and by A_i the $m \times (n-1)$ matrix obtained from A by omission of a_i, the ith LLS estimate y_i of the linear relations between the columns of A is obtained by minimizing

$$\|a_i - A_i y_i\| \tag{2}$$

where $\|\cdot\|$ is the standard 2-norm of a vector. Because $r(A) = n$, it follows that $r(A_i) = n - 1$. Hence, the vector y_i that minimizes (2) is unique and given by

$$y_i = (A_i^t A_i)^{-1} A_i^t a_i. \tag{3}$$

Obviously, there are n such least squares solutions, for $i = 1, \cdots, n$. The geometrical interpretation for each of the LLS solutions is well known and can easily be derived from the singular value decomposition (SVD) of A_i [7]:

$$A_i = U_i D_i V_i^t \tag{4}$$

where, in this case, U_i is $m \times (n-1)$ orthonormal, $D_i\,(n-1) \times (n-1)$ diagonal with positive diagonal elements (the singular values), and $V_i (n-1) \times (n-1)$ orthonormal. Then (3) can be rewritten as

$$y_i = V_i D_i^{-1} U_i^t a_i$$

and the residual vector can be decomposed as

$$\tilde{a}_i = a_i - A_i y_i = (I_n - U_i U_i^t) a_i \tag{5}$$

which is nothing else than the orthogonalization of a_i onto the range of A_i. The modification of the data matrix A for the ith LLS solution hence consists of an orthogonal decomposition of the ith column a_i into its orthogonal projection onto the range of A_i and the residual vector given by (5). This is of course nothing else than the well-known *orthogonality principle*, which is sometimes exploited to derive elegant solutions to least squares problems [12]. The conclusion is that *each of the n LLS solution leads to the modification of only 1 column vector of the data matrix A in order to make it singular.*

It is less known that the n LLS solutions y_i, as in (3), can be derived from the columns of the inverse of the sample covariance matrix Σ of the data

$$\Sigma = A^t A.$$

Without loss of generality, we derive this result for $i = n$. The sample covariance matrix can then be partitioned as

$$\Sigma = (A_n \quad a_n)^t (A_n \quad a_n) = \begin{pmatrix} A_n^t A_n & A_n^t a_n \\ a_n^t A_n & a_n^t a_n \end{pmatrix}.$$

Via a well-known lemma for the inverse of a partitioned matrix, it follows that

$$\Sigma^{-1} = \begin{pmatrix} (A_n^t A_n - A_n^t a_n (a_n^t a_n)^{-1} a_n^t A_n)^{-1} & -(A_n^t A_n)^{-1} A_n^t a_n / \alpha_n \\ -a_n^t A_n (A_n^t A_n)^{-1} / \alpha_n & 1/\alpha_n \end{pmatrix}$$

where

$$\alpha_n = a_n^t a_n - a_n^t A_n (A_n^t A_n)^{-1} A_n^t a_n. \tag{6}$$

The scalar α_n has a double significance as follows.

- When the nth column of Σ^{-1} is multiplied with $-\alpha_n$, the vector consisting of the first $n-1$ components is the same as the nth LLS solution y_n derived in (3).
- Using SVD of A_n defined in (4) results in

$$\alpha_n = a_n^t a_n - a_n^t U_n U_n^t a_n$$
$$= a_n^t (I_n - U_n U_n^t)^2 a_n.$$

Hence, α_n is nothing else than the square of the norm of the residual vector $\tilde{a}_n$ (5) from the nth least squares solution.

These results apply for the other columns of Σ^{-1} and the corresponding diagonal elements in a trivial manner.

III. Total Linear Least Squares

For each of the LLS solutions, only one column of the data matrix A is modified. However, the fact that all columns of A are noisy suggests an estimation scheme in which all columns are modified in order to make the data matrix singular. This is achieved by the total linear least squares solution (TLLS). It looks for the matrix $\hat{A}$ with $r(\hat{A}) = n - 1$ that is closest to A in Frobenius norm:

$$\min_{\text{over all } \hat{A},\ \text{rank}(\hat{A}) = n} - 1 \|A - \hat{A}\|_F^2.$$

The $n \times 1$ vector x satisfying $\hat{A}x = 0$ is then called the TLLS solution. The solution is immediate from the singular value decomposition of A:

$$A = UDV^t = \sum_{i=1}^{n} u_i \delta_i v_i^t.$$

In case $\delta_{n-1} > \delta_n$ (the generic case), the optimal solution $\hat{A}$ is given by the first $n-1$ terms

$$\hat{A} = \sum_{i=1}^{n-1} u_i \delta_i v_i^t$$

and the TLLS solution x is up to a scalar

$$x_{TLLS} = v_n. \tag{7}$$

For a proof, see [6]. The nongeneric case occurs whenever $\delta_{n-1} = \delta_n$ and is analyzed in detail in [13] but will not be considered here.

The geometrical interpretation follows immediately from the singular value decomposition of A

$$A = \sum_{i=1}^{n} u_i \delta_i v_i' = \sum_{i=1}^{n-1} u_i \delta_i v_i' + u_n \delta_n v_n' = \hat{A} + \tilde{A}.$$

The TLLS estimation scheme decomposes the data in a model for the exact part, $\hat{A}$, and one for the pure noise part, $\tilde{A}$, with $\text{rank}(\tilde{A}) = 1$. Hence, all columns of $\tilde{A}$ are proportional to the nth left singular vector u_n and orthogonal to the range of $\hat{A}$. Despite the fact that, contrary to the LLS solution, now all columns of A are modified, the solution remains very structured: the noise model is a rank one matrix!

Using the SVD of A, it is a straightforward exercise to show that the TLLS solution x_{TLLS} is nothing else than the eigenvector of the sample covariance matrix $\Sigma = A'A$, corresponding to the smallest eigenvalue.

IV. Rank One Modifications

Both the LLS and TLLS solutions decompose the matrix A as

$$A = \hat{A} + \tilde{A}$$

with the additional properties

$$\text{rank}(\tilde{A}) = 1 \tag{8}$$

$$\tilde{A}'\hat{A} = 0. \tag{9}$$

We shall first investigate the properties of rank one matrices $\tilde{A}$ that lower the rank of A (condition 8) and use the results to characterize the rank one matrices $\tilde{A}$ that lower the rank of A and are orthogonal to the resulting $\hat{A}$ (conditions 8 *and* 9).

Lemma 1—Rank Reduction of Rank One: Let A be an $m \times n$ matrix, $m > n$ and $\text{rank}(A) = n$. Let $\tilde{A} = p\sigma q'$ be a rank one matrix with σ a nonzero scalar, p an $m \times 1$ and q an $n \times 1$ vector with $\|p\| = \|q\| = 1$. If:

$$(A - p\sigma q')x = 0 \tag{10}$$

then

1) p belongs to the range of A and

$$p'A(A'A)^{-1}A'p = 1, \tag{11}$$

2) the solution x does not depend on q

$$x = (A'A)^{-1}A'p\beta \tag{12}$$

where β is a nonzero scalar; and

3) σ is determined by p and q

$$\sigma^{-1} = q'(A'A)^{-1}A'p. \tag{13}$$

Proof: Observe that from $\text{rank}(A) = n$ it follows that $q'x \neq 0$. From (10) one finds that $p = Ax/(\sigma(q'x))$ which proves the first property and $p'A(A'A)^{-1}A'p = 1$ holds for any unit vector p in the range of A. Put $\beta = \sigma(q'x)$. Premultiplication of p with $(A'A)^{-1}A'$ proves (12). In order to prove (13), substitute (12) in (10) to find that $A(A'A)^{-1}A'p\beta = p\sigma q'(A'A)^{-1}A'p\beta$. Premultiplication with p' then proves (13). □

If, in addition to the conditions of Lemma 1, we also require that $\hat{A}'\tilde{A} = 0$, the following result applies.

Theorem 1: Rank reduction of rank one with orthogonality requirements. Consider $\hat{A}$, $\tilde{A}$ as in Lemma 1 with the additional requirement that

$$\hat{A}'\tilde{A} = 0. \tag{14}$$

Then

1)
$$A'p = q\sigma \tag{15}$$

2)
$$\sigma^{-2} = q'(A'A)^{-1}q \tag{16}$$

3)
$$\sigma^2 = p'AA'p \tag{17}$$

4)
$$p = A(A'A)^{-1}q\sigma. \tag{18}$$

Proof: Equation (15) follows immediately from (14) because $\hat{A}'\tilde{A} = (A' - q\sigma p')p\sigma q' = 0 = (A'p - q\sigma)q'$. Because $q \neq 0$, $A'p = q\sigma$. Equation (16) follows from a combination of (10)–(12) and (15). Equation (17) follows from premultiplying (15) with $p'A$. Finally, (18) follows from (10) and (11) combined with (15). □

Similar results can also be derived via the sample covariance matrix. Because of the orthogonality property (14) and the additivity of the noise, the sample covariance matrix can be written as

$$\begin{aligned}\Sigma &= A'A \\ &= \hat{A}'\hat{A} + \tilde{A}'\tilde{A} \\ &= \hat{\Sigma} + \tilde{\Sigma}\end{aligned}$$

where $\text{rank}(\tilde{\Sigma}) = 1$. While Σ is positive definite, both $\hat{\Sigma}$ and $\tilde{\Sigma}$ are nonnegative definite. The following theorem characterizes all nonnegative definite rank one matrices $\tilde{\Sigma}$ that, when subtracted from Σ, result in a nonnegative definite matrix $\hat{\Sigma}$ of rank $n-1$.

Theorem 2: Let Σ be an $n \times n$ positive definite matrix. Let $\tilde{\Sigma}_q = q\sigma_q^2 q'$ be a rank one matrix, with q an $n \times 1$ vector with $\|q\| = 1$ and σ_q a nonzero scalar. Define $\hat{\Sigma} = \Sigma - \tilde{\Sigma}_q$. Then the conditions

- $\text{rank}(\hat{\Sigma}) = n - 1$
- $\hat{\Sigma}$ is nonnegative definite

are satisfied if and only if

$$\sigma_q^{-2} = q'\Sigma^{-1}q. \tag{19}$$

The solution x_q of $(\Sigma - \tilde{\Sigma}_q)x_q = 0$ is given by

$$x_q = \Sigma^{-1}q. \tag{20}$$

Proof:

If Part: With $\tilde{\Sigma}_q$, σ_q, and x_q as above, it follows that $(\Sigma - \tilde{\Sigma}_q)x_q = \Sigma\Sigma^{-1}q - q\sigma_q^2 q'\Sigma^{-1}q = 0$. Because $\text{rank}(\Sigma) = n$ and $\text{rank}(\tilde{\Sigma}_q) = 1$, it follows that $\text{rank}(\Sigma - \tilde{\Sigma}_q) = n - 1$. Moreover, the matrix $(\Sigma - \tilde{\Sigma}_q)\Sigma^{-1}(\Sigma - \tilde{\Sigma}_q)$ is obviously nonnegative definite. But

$$\begin{aligned}(\Sigma - \tilde{\Sigma}_q)\Sigma^{-1}(\Sigma - \tilde{\Sigma}_q) &= \Sigma - 2\tilde{\Sigma}_q + \tilde{\Sigma}_q\Sigma^{-1}\tilde{\Sigma}_q \\ &= \Sigma - 2\tilde{\Sigma}_q + q\sigma_q^2 q' \\ &= \Sigma - \tilde{\Sigma}_q.\end{aligned}$$

Hence, $\Sigma - \tilde{\Sigma}_q$ is nonnegative definite.

Only If Part: The conditions $(\Sigma - q\sigma_q^2 q')x_q = 0$ and $\text{rank}(\Sigma) = n$ imply that $q'x_q \neq 0$ and that $x_q = \Sigma^{-1}q\sigma_q^2 q'x_q$. Premultiplication with q' results in $q'x_q = q'\Sigma^{-1}q\sigma_q^2 q'x_q$ from which it follows that $\sigma_q^{-2} = q'\Sigma^{-1}q$. It is straightforward to show that $x_q = \Sigma^{-1}q$ is the only solution for a fixed q. □

Theorems 1 and 2 characterize all rank one modifications of a given matrix A as $\hat{A} = A - \tilde{A}$ where $\text{rank}(\tilde{A}) = 1$ and $\hat{A}'\tilde{A} = 0$. It is the choice of the vector q that fixes the solution x_q [via (20)], the noise "energy" [via (19)] and the direction p in which all of the columns of A will be modified to make it singular [via (18)]. The n LLS solutions and the TLLS solution correspond to certain specific choices of the vector q which result in special geometrical, numerical, and statistical properties.

The ith LLS solution and its properties, as explored in Section II, follow from the choice of $q = e_i$ where e_i is the unit vector with the ith component 1 and all others zero. From (16) and (19), we find that $\sigma_{e_i}^{-2} = e_i'(A'A)^{-1}e_i = e_i'\Sigma^{-1}e_i$. Hence, the ith diagonal element of Σ^{-1} is the inverse of the square of the norm of the residual as in (6). From (18), $p = A(A'A)^{-1}e_i\sigma_{e_i}$, hence, from (12), $x_{LLS_i} = (A'A)^{-1}e_i(\sigma_{e_i}\beta)$, which is (20), up to a scalar. The ith LLS solution hence follows from the ith column of Σ^{-1}, which was also found in Section II. The noise covariance matrix $\tilde{\Sigma}_{e_i}$, for the ith LLS solution is equal to $\tilde{\Sigma}_{e_i} = e_i\sigma_{e_i}^2 e_i'$, hence contains only one nonzero element, which is the inverse of the ith diagonal element of Σ^{-1}.

The TLLS solution is found by choosing $q = v_n$, the eigenvector of Σ corresponding to the smallest eigenvalue. The corresponding $\sigma_{v_n}^{-2}$ follows from (19) and $\sigma_{v_n}^2$ is the smallest eigenvalue of Σ. Hence, of all possible rank one modifications in Theorem 2, the TLLS modification is the one with the smallest Frobenius norm. From (18) we get that $p = u_n$, the right singular vector corresponding to the smallest singular value. The rank one noise covariance is given by

$$\tilde{\Sigma}_{v_n} = v_n \sigma_{v_n}^2 v_n^t. \quad (21)$$

From (20), the TLLS solution is given by

$$x_{TLLS} = \Sigma^{-1} v_n = v_n \sigma_{v_n}^{-2}. \quad (22)$$

This last relation establishes an interesting relation with the results reported in [2], [3], and [9]. Since the columns of Σ^{-1} are (up to a scalar) the LLS solutions, it follows that x_{TLLS} is a nonnegative combination (=linear combination with nonnegative weights) *of the LLS solutions, if v_n has nonnegative components*. A sufficient condition for this is that Σ^{-1} is sign-similar to an elementwise positive matrix (i.e., by symmetric sign changes of the form $S\Sigma^{-1}S$, where S is a diagonal sign matrix, it can be made elementwise positive). In this case, the Perron–Frobenius theorem for nonnegative matrices applies. Since v_n is the eigenvector corresponding to the largest eigenvalue of Σ^{-1}, its components can be chosen to be nonnegative if Σ^{-1} is sign-similar to an elementwise positive matrix. The fact that Σ^{-1} is (sign-similar to) an elementwise positive matrix also implies that the LLS solutions can all be transformed to the positive orthant by appropriate sign changes. Assume that we start from an exact $m \times n$ matrix $\hat{A}$ of rank$(\hat{A}) = n - 1$ (which is well conditioned in the sense that its two smallest nonzero singular values are not too close). Then for sufficiently small noise matrices $\tilde{A}$ (not necessarily of rank one), added to $\hat{A}$, the inverse of the sample covariance matrix will be (sign similar to) an elementwise positive matrix, the columns of which are the least squares solutions. In this case, the TLLS solution is a nonnegative combination of these least squares solutions, which will be closer to the orthogonal "exact" solution of $\hat{A}\hat{x} = 0$ than each of the least squares solutions. It is proved in [2], [3], and [9] that, if Σ^{-1} is sign-similar to an elementwise positive matrix, the LLS solutions are the vertices of a simplex with the following property: for every vector x within the simplex, there exists a nonnegative diagonal matrix $\tilde{\Sigma}$ with $\Sigma x = \tilde{\Sigma} x$ such that $\Sigma - \tilde{\Sigma}$ is nonnegative definite of rank $n - 1$. Hence, (22) expresses the fact that the TLLS solution also lies within this simplex. While the corresponding covariance matrix (21) is not diagonal, it is possible to determine a nonnegative diagonal matrix $\tilde{\Sigma}$ such that $(\Sigma - \tilde{\Sigma})x_{TLLS} = 0$ with $\Sigma - \tilde{\Sigma}$ nonnegative definite of rank $n-1$: simply choose $\tilde{\Sigma} = \sigma_{v_n}^2 I_n$. The qualitative conditions under which Σ^{-1} will be sign-similar to an elementwise positive matrix are precisely those stated in the Introduction. If, for instance, the Frobenius norm of the added noise matrix $\tilde{A}$ is increased, the (positive) simplex of the LLS solutions grows "bigger" until its vertices reach the orthant plants bordering the positive orthant. At this point, the matrix Σ^{-1} ceases to be (sign-similar to) an elementwise positive matrix. Further research is, however, needed for a more quantitative description.

References

[1] T. W. Anderson, "The 1982 Wald memorial lectures: Estimating linear statistical relationships," *Ann. Stat.*, vol. 12, no. 1, pp. 1-45, 1984.

[2] B. De Moor and J. Vandewalle, "A geometrical approach to the maximal corank problem in the analysis of linear relations," in *Proc. 25th IEEE-CDC*, Athens, Greece, Dec. 1986, pp. 1990–1995.

[3] —, "The uniqueness versus the non-uniqueness principle in the identification of linear relations from noisy data," in *Proc. 25th IEEE-CDC*, Athens, Greece, Dec. 1986, pp. 1663–1665.

[4] B. De Moor, "Mathematical concepts and techniques for the analysis of static and dynamic systems," Ph.D. dissertation, Dep. Elec. Eng., Katholieke Universiteit Leuven, Belgium, June 1988.

[5] G. H. Golub, "Some modified matrix eigenvalue problems," *SIAM Rev.*, vol. 15, no. 2, pp. 318-334, Apr. 1973.

[6] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883-893, Dec. 1980.

[7] —, *Matrix Computations.* North Oxford Academic, Johns Hopkins University Press, 1983.

[8] T. Kailath, "Linear least squares estimation," in *Benchmark Papers in Electrical Engineering and Computer Science*, vol. 17. Stroudsburg, PA: Dowden, Hutchinson & Ross, 1977.

[9] R. E. Kalman, "System identification from noisy data," in *Dynamical Systems II*, A. R. Bednarek and L. Cesari, Eds. New York: Academic, 1982.

[10] C. Lawson and R. Hanson, *Solving Least Squares Problems*, Prentice-Hall Series in Automatic Computation. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[11] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: M.I.T. Press, 1983.

[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill Series in Electrical Engineering. New York: McGraw-Hill, 1984.

[13] S. Van Huffel, "Analysis of the total least squares problem and its use in parameter estimation," Ph.D. dissertation, Dep. Elec. Eng., Katholieke Universiteit Leuven, Belgium, June 1987.

A Hamiltonian–Jacobi Algorithm

RALPH BYERS

Abstract—**This note adapts the nonsymmetric Jacobi iteration to the special structure of Hamiltonian matrices. This Hamiltonian–Jacobi algorithm uses symplectic-unitary similarity transformations to solve algebraic Riccati equations through the Hamiltonian–Schur form. It preserves Hamiltonian structure without using a condensed form. Although it converges too slowly for use on conventional serial computers, it may be attractive for some highly parallel architectures.**

I. Introduction

A matrix $M \in C^{2n \times 2n}$ is *Hamiltonian* if it is of the form

$$M = \begin{bmatrix} A & G \\ F & -A^H \end{bmatrix} \quad (1)$$

where $A, F, G \in C^{n \times n}$, $F = F^H$ $G = G^H$ [6], [11]. The superscript H denotes complex conjugate transpose. Solutions $X \in C^{n \times n}$ to the algebraic Riccati equation

$$F + A^H X + XA - XGX = 0 \quad (2)$$

correspond to invariant subspaces of M. If X satisfies (2), then

$$\begin{bmatrix} A & G \\ F & -A^H \end{bmatrix} \begin{bmatrix} I_n \\ -X \end{bmatrix} = \begin{bmatrix} I_n \\ -X \end{bmatrix} [A - GX].$$

In particular, the columns of $\begin{bmatrix} I_n \\ -X \end{bmatrix}$ span an n-dimensional invariant subspace of M. Conversely, if $Y \in C^{n \times n}$ is nonsingular and $Z \in C^{n \times n}$ is such that the columns of $\begin{bmatrix} Y \\ Z \end{bmatrix}$ span an n-dimensional invariant subspace of M, then $X = -ZY^{-1}$ satisfies (2). As it arises in control theory, the desired solution is stabilizing in the sense that all eigenvalues of $A - GX$ have negative real part. This implies that the associated n-dimensional invariant subspace is the one corresponding to the eigenvalues of M with negative real part. Under mild assumptions, such an invariant subspace exists and is unique [10].

Numerical methods for finding the invariant subspace through an eigenvalue-eigenvector factorization of M were proposed by MacFarlane [13] and Potter [17]. Laub [12] improved numerical stability and lowered work requirements by using a Schur decomposition to find the subspace.

Manuscript received January 20, 1989; revised August 18, 1989. Some of this work was completed while the author was employed in the Mathematics Department, North Carolina State University, Raleigh, NC. This work was supported in part by the National Science Foundation under Grant CCR-8820882, and University of Kansas General Research Allocations 3758-20-038 and 3692-20-038.

The author is with the Department of Mathematics, University of Kansas, Lawrence, KS 66045.

IEEE Log Number 9034491.